



Leckie, G. (2018). Monitoring School Performance Using Value-Added and Value-Table Models: Lessons from the UK. In *Studies in Theoretical and Applied Statistics: Proceedings of the 48th Scientific meeting of the Italian Statistical Society, Salerno, Italy, June 8-10* (pp. 251-260). (Springer Proceedings in Mathematics & Statistics; Vol. 227). https://doi.org/10.1007/978-3-319-73906-9_23

Peer reviewed version

License (if available):
Unspecified

Link to published version (if available):
[10.1007/978-3-319-73906-9_23](https://doi.org/10.1007/978-3-319-73906-9_23)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Springer at https://link.springer.com/chapter/10.1007%2F978-3-319-73906-9_23 . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Monitoring school performance using value-added and value-table models: Lessons from the UK

George Leckie

Abstract Since 1992, the UK Government has published so-called ‘school league tables’ summarizing the average attainment and progress made by pupils in each state-funded secondary school in England. In this article, we statistically critique and compare prominent past, current and forthcoming value-added and value-table measures of school performance. We discuss the advantages and disadvantages of the different measures as well as their underlying statistical models.

Abstract *Abstract in Italian*

Key words: value-added model, value-table, school accountability

1 Introduction

The UK has a long history of publishing ‘school league tables’ summarising pupils’ examination and test results. Over time, increasingly sophisticated measures have been introduced culminating in 2006 with contextual value-added (CVA), a multilevel modelling ‘value-added’ based approach. However, in 2011 the Government withdrew CVA replacing it with expected progress (EP), a simpler ‘value-table’ approach. In this paper we: question the Government’s reasons for withdrawing CVA; we argue that EP suffers from serious design flaws; and we show that CVA and EP lead to very different rankings and therefore that choice of school performance measure has very important ramifications for school accountability.

¹ George Leckie, Centre for Multilevel Modelling and Graduate School of Education, University of Bristol, UK; email: g.leckie@bristol.ac.uk

2 Background to national tests, school performance measures and school league tables in England

The English education system consists of a primary phase of education (ages 4–11, years R–6) followed by a secondary phase of education (ages 11–16, years 7–11). Effectively all pupils change schools at the transition between the two phases. At the end of primary schooling, all pupils sit national Key Stage 2 (KS2) tests in English and maths. These are measured using continuous point scores, but are discretised into levels for reporting purposes: W (working below level 1), 1, 2, 3, 4, 5. At the end of secondary schooling, all pupils sit national GCSE examinations in English, mathematics as well as in a range of other subjects of their choosing. Attainment in each subject is measured using GCSE grades: U, G, F, E, D, C, B, A, A*. School league tables are then constructed summarising schools' performances in these KS2 tests and GCSE examinations (<http://www.education.gov.uk/schools/performance/>). Our focus is on secondary school performance measures and therefore schools' GCSE performances, and especially the average progress made by pupils in each school during secondary schooling between their KS2 tests and their GCSE examinations.

The Government gives three main justifications for publishing school league tables. First school league tables are published to support parental school choice based on schools' ability to teach the national curriculum, and to therefore create competition and a free market in education (Education Reform act 1988). An important element of this first argument is that the tables are routinely republished by the media and so have a very high national profile. Second, they are published to enable school accountability; publically funded schools should be held publically accountable. Indeed, Ofsted, the official schools' inspectorate system in part choose which schools to inspect on the basis of schools' league table results. Schools whose results do not improve face takeover by neighbouring schools or ultimately closure. Third, they are published to promote school improvement via school self-reflection and the identification of effective practices being employed in successful schools. Indeed, a number of commercial and charitable companies now sell to schools pupil performance monitoring software and other services based on the same data which underlies the Government's tables.

An important distinction to be made is between 'attainment' and 'progress' school performance measures. Attainment measures aim to report the average 'status' of pupils at the end of secondary schooling. The headline attainment measure in England for effectively the last 20 years has been the percentage of pupils achieving 5 or more A*-C GCSE grades (5 A*-C). Attainment measures may give useful information regarding school inequalities, but it is crucial to realise that in England and other education systems more generally they reflect differences in school intake composition more than school processes. In contrast, progress measures (e.g., CVA and EP) aim to report the average 'growth' or 'improvement' made by pupils during secondary schooling. Progress measures are generally considered the fairer and more meaningful way to measure school performance for school choice, accountability and improvement purposes.

Table 1 reports various performance measures for schools in Bristol in 2010 and should be referred back to as we discuss the various measures in more detail below.

Table 1: City of Bristol 2010 school league table.

School	n	5 A*-C	CVA	CVA lower	CVA upper	EP English	EP maths
Ashton Park School	180	49	1003	994	1013	66	70
Bedminster Down School	191	40	989	979	998	74	48
Bridge Learning Campus - Secondary	145	34	1003	993	1014	64	44
Brislington Enterprise College	216	37	970	962	979	60	40
Bristol Brunel Academy	158	45	1005	994	1016	69	62
Bristol Cathedral Choir School	75	75	1002	987	1017	95	77
Bristol Metropolitan Academy	127	39	1011	999	1023	76	61
The City Academy Bristol	183	36	1036	1027	1046	71	49
Colston's Girls' School	68	91	1010	992	1027	100	90
Cotham School	180	77	1016	1006	1026	86	85
Fairfield High School	194	49	1004	994	1014	73	63
Henbury School	161	39	1001	991	1011	66	54
Merchants' Academy	124	25	1010	998	1021	56	26
Oasis Academy Brightstowe	93	29	1028	1015	1041	62	37
Oasis Academy Bristol	115	29	1007	995	1019	56	36
Orchard School	172	37	1005	995	1015	69	51
St Bede's Catholic College	185	72	1006	996	1016	80	71
St Bernadette Catholic Secondary School	152	37	980	969	990	66	47
St Mary Redcliffe and Temple School	207	70	1013	1004	1022	86	75

Notes: Table reproduced from http://www.education.gov.uk/schools/performance/archive/schools_10/pdf_10/801.pdf. n = number of pupils at the end of GCSE; 5 A*-C = Percentage of pupils with five or more GCSEs (or equivalent qualifications) at grade A* to C; CVA = Contextual value-added score (national average = 1000); CVA lower = Lower limit of CVA 95% confidence interval; CVA upper = Upper limit of CVA 95% confidence interval; EP English = Percentage of pupils making expected progress in English; EP Maths = Percentage of pupils making expected progress in mathematics.

3 Contextual value-added (2006-2010)

The Government's CVA measure is based on the standard approach to modelling value-added in the school-effectiveness literature which is to fit a two-level pupil-within-schools random-intercept model to pupils' final attainment adjusting for pupil prior attainment and other pupil socioeconomic and demographic characteristics (Goldstein, 2011). The reported CVA scores are simply the empirical Bayes predicted school random effects. These scores are presented with 95% confidence intervals to communicate their statistical uncertainty. Conceptually, CVA scores and value-added

scores more generally can be viewed as school-level averages of the differences between pupils' actual and predicted GCSE scores. A simplified version of the CVA model can be written as

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 z_{ij} + u_j + e_{ij} \quad (1)$$

where

- y_{ij} denotes the GCSE score of pupil i ($i = 1, \dots, n_j$) in school j ($j = 1, \dots, J$)
- x_{ij} denotes their KS2 score
- z_{ij} denotes their free school mean status
- u_j denotes their school's value-added effect or CVA score
- e_{ij} denotes the pupil residual

The full model enters KS2 score as a flexible polynomial and includes a much wider range of pupil socioeconomic and demographic characteristics including pupil age, gender, ethnicity, special education needs status and residential deprivation score. The GCSE score is summed over pupils' best eight GCSE results, while their KS2 score is averaged across their separate results for English and mathematics.

3.1 *The Government's justifications for withdrawing CVA*

The government withdrew CVA in 2010 citing a number of justifications (DfE, 2010). First they argued that '[CVA] is difficult for the public to understand'. Clearly CVA is more complex than simply reporting school average exam scores. However, the notion of making adjustments for differences in schools' student compositions in terms of their prior attainment and other factors isn't in itself intrinsically difficult to understand. There is no need for the public to understand the statistical details of the model in order to interpret the adjusted school-mean scores. Perhaps the real problem is that the Government did not do enough to explain and communicate CVA? For example, one had to delve deep into the technical documentation to find out what the CVA unit of measurement was. Clearly the notion of 95% confidence intervals is also hard for the public to understand, however perhaps the Government should have explored various graphical approaches for communicating statistical uncertainty rather than simply reporting the confidence intervals in tabular form (Leckie and Goldstein, 2011; Leckie et al., 2016). It is also worth noting that the methodology underlying CVA is the same as that used in Hong Kong (Schools Value-added Information System, SVAIS) and simpler than that underlying other school performance measures published around the world (e.g., Tennessee's e Value-Added Assessment System, TVAAS, or Australia's similar schools methodology).

The Government's second reason for ending CVA was that 'recent research shows [CVA] to be a less strong predictor of success than raw attainment measures'. It is not

entirely clear what the Government are trying to say here (they don't cite the research they refer to). It sounds like they are saying that a pupil's GCSE score is more strongly predicted by their KS2 score than by their school's CVA score. However, predicting GCSE success was never the aim of CVA; the aim was to measure the effects schools actually had on their pupils.

The Government's third reason for ending CVA was that '[CVA] also has the effect of expecting different levels of progress from different groups of pupils on the basis of their ethnic background, or family circumstances, which we think is wrong in principle'. However, CVA did not *a priori* expect different levels of progress from different pupil groups, rather it adjusted for such differences if they arose. The reality is that some pupil groups *do* make less progress than others and that this must be adjusted for if we are to make fair comparisons between schools. Failure to do so leads to 'comparing apples and oranges'

Expanding on this theme, the Government argue that 'It is morally wrong to have an attainment measure which entrenches low aspirations for children because of their background'. The Government are arguing that by adjusting for pupil background, CVA led to a system-level acceptance that socially and other disadvantaged pupil groups will make less progress than their more advantage peers. Although not stated explicitly, the real concern appears to be that some schools started to use the published CVA model to set differential GCSE targets for current pupils based on their background. This was never the purpose of CVA and reflects the perverse incentives that so often arise with high-stakes school league tables.

Finally, the Government argued that 'We should expect every child to succeed and measure schools on how much value they add for all pupils, not rank them on the ethnic make-up of their intake...'. This statement suggests a fundamental misunderstanding. CVA explicitly adjusted for as many of the observed differences between schools' intakes as possible in order to remove their influence from schools' rankings. In contrast, it is when one ignores these differences that one implicitly ranks schools on the make-up of their intakes.

4 Expected progress (2011-2015)

The Government's Expected Progress (EP) measures is based on value-table methodology (Castellano and Ho, 2013). EP is published separately for English and mathematics. EP is calculated simply as the percentage of pupils making three levels of progress between KS2 and GCSE; it ignores pupils' socioeconomic and demographic characteristics. The Government's introduction of EP can be seen as an explicit attempt to address the flaws they perceived in CVA. Specifically, EP is designed to be both easy for the public to understand and blind to all differences between schools' intakes other than their prior attainment.

Every pupil is effectively set a target GCSE grade in English and separately in mathematics as a function of their KS2 levels in those subjects. Table 2 presents this idea in tabular form. Thus, for example, low prior attainers (those who achieved KS2 level 3) are expected to achieve a D GCSE grade or higher, while middle prior attainers (KS2 level 4) are expected to achieve a C or higher, and high prior attainers (KS2 level 5) are expected to achieve a B or higher. Essentially, all pupils are expected to progress by 3 (or more) levels during the five year duration of secondary schooling.

Table 2: Table showing how expected progress in English and mathematics is calculated.

KS2 level	GCSE target grade / level									
	?	U	G	F	E	D	C	B	A	A*
?	No	No	?	?	?	?	?	Yes	Yes	Yes
W	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
3	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
4	No	No	No	No	No	No	Yes	Yes	Yes	Yes
5	No	No	No	No	No	No	No	Yes	Yes	Yes

Notes: W = Working towards level 1; ? = No result; No = EP not made; Yes = EP made. Table reproduced from Department for Education (2015a).

We can write this value-table model for English (or equally for mathematics) as

$$EP_{ij} = I(y_{ij} - x_{ij} \geq 3), \quad \overline{EP}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} EP_{ij} \quad (2)$$

where

- y_{ij} denotes the GCSE level associated with the English grade of pupil i ($i = 1, \dots, n_j$) in school j ($j = 1, \dots, J$)
- x_{ij} denotes their English KS2 level
- EP_{ij} denotes whether they made expected progress in English (i.e., 3 or more levels of progress)
- \overline{EP}_j denotes the school proportion of pupils making expected progress in English

An important aspect of EP is its role in national ‘floor standards’ introduced by the Government also in 2011. A school is judged ‘underperforming’ if less than 40% of pupils achieve 5 A*-C, but is exempted if they are in the top half nationally in EP in both English and mathematics. Thus, EP is meant to play a central balancing role in these judgements. Schools judged underperforming face increased scrutiny from Ofsted, potential takeover by neighbouring schools, or even closure.

4.1 Statistical concerns with expected progress

Our first statistical concern with EP is that it will bring about perverse incentives whereby schools concentrate their efforts on those pupils who are borderline in terms of potentially making EP (i.e., those pupils operating just below the No/Yes boundary in Table 2). This perverse incentive is largely driven by the fact that the transition values of the value table are binary (EP is a threshold measure). There are no partial rewards for just missing target grades; no additional rewards for surpassing target grades.

Our second statistical concern with EP is that, nationally, there is a strong dependency on prior attainment. Figure 1 shows that the national percentage of pupils making EP increases sharply with KS2 level (we have restricted the plot to KS2 levels 3, 4 and 5 as these account for the vast majority of pupils, over 95%). Thus, it is harder for low prior attainers to make expected progress than it is for high prior attainers. Low prior attainers are set relatively tough target GCSE grades while high prior attainers are set relatively easy target GCSE grades. Schools with higher prior attaining intakes will therefore do better on EP. EP therefore under-adjusts for school differences in prior attainment. EP is not a pure measure of progress in the way that CVA is.

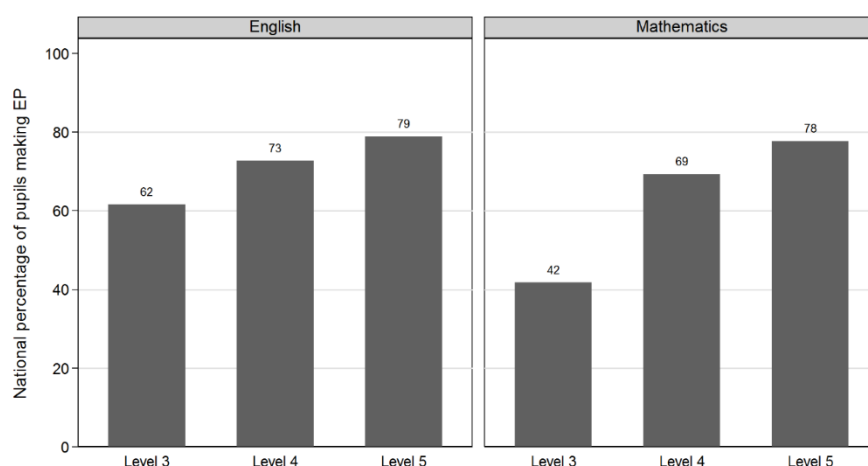


Figure 1: National percentage of pupils making EP (expected progress) during secondary schooling against KS2 levels (3, 4 and 5) in 2014, reported separately for English and mathematics.

A related concern reveals itself when we inspect the relationship between the national percentage making EP and the underlying continuous prior attainment score in each subject. Figure 2 plots a scatterplot of this relationship where the size of the plotted points is proportional to the number of students with each KS2 score. First note that the overall positive association between the percentage of pupils making EP and their prior attainment is revealed to be even stronger than before. For example, in maths, the percentage of pupils making their target GCSE grade ranges from below 20% to above 80% as we move from the lowest to the highest KS2 scores. However, in

contrast to Figure 1, we now also see that EP has an illogical sawtooth (zig zag) dependency on prior attainment with sharp discontinuities in the probability of making EP as we move from the top of one KS2 level to the bottom of the next. Thus, with this approach, pupils with the same prior attainment are set very different educational challenges in terms of their target grades. This is clearly undesirable.

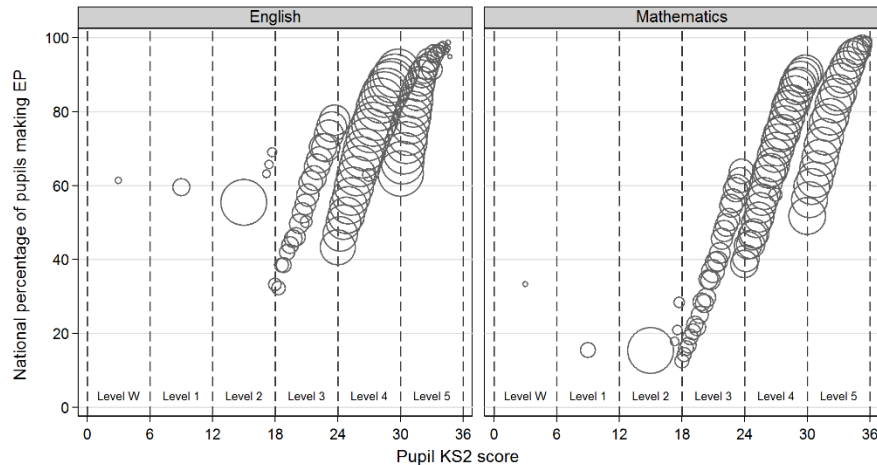


Figure 2: National percentage of pupils making EP (expected progress) during secondary schooling against KS2 score in 2014, reported separately for English and mathematics. The magnitude of the hollow circles are proportional to the national number of pupils with that KS2 score. The dashed vertical lines denote the KS2 level thresholds. Level W = working towards level 1. For clarity, the plot is restricted to values of KS2 score for which there were at least 100 pupils nationally.

Our third statistical concern is that EP takes no account of pupils' socioeconomic and demographic characteristics and therefore will be biased in favour of schools which serve more advantaged pupil groups.

Our fourth statistical concern with EP is that it makes no attempt to quantify and communicate the statistical uncertainty in measuring school effects. There is no obvious way for users to establish whether measured differences between schools, or differences from national averages and floor standards, are meaningful, or whether they more likely reflect the variations of chance. Consider a school with 180 pupils where 70% make EP. The associated 95% Wald binomial confidence interval ranges from 63% to 77% and so the school has a ± 7 percentage point margin of error which would be completely unacceptable in any survey or poll of public opinion. When we plot the 95% confidence interval for every school in the country (Figure 3), we see that over a third of schools cannot be distinguished from the national average in either English or mathematics.

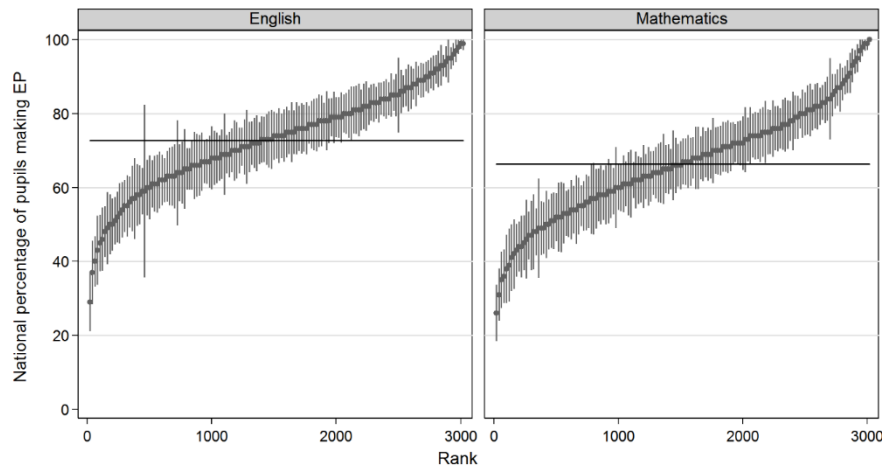


Figure 3: EP (expected progress) scores in 2014 with 95% Wald binomial confidence intervals presented in rank order of magnitude, reported separately for English and mathematics. Higher ranks denote higher performances. The horizontal lines denote the national average EP scores. The confidence intervals are approximate hence the upper bounds exceeding a value of 100 for a minority of schools with exceptionally high EP scores. For clarity, the plot shows every 20th school.

5 Expected progress vs. Contextual value-added

We have explained how CVA and EP, based on value-added and value-table methodologies, are fundamentally different measures of school progress. However, if the two measures lead to similar rankings then it could be argued that our arguments are largely academic. In this section we therefore analyse the 2010 data (3,056 schools) to contrast the two methods of calculating school progress empirically.

Table 3 reports Pearson correlations between the CVA, EP, 5 A*-C and KS2 APS (average point score across English and maths). We see that CVA and EP are only moderately positively correlated (correlations of 0.36 and 0.29 between CVA and EP in English and maths). EP is much more highly correlated with 5 A*-C (correlations of 0.85 and 0.89) and is therefore closer to being a pure attainment measure of school performance than a pure progress measure. This is supported by the high correlations between EP & KS2 APS (correlations of 0.64 and 0.67), whereas there is effectively no relationship between CVA & KS2 APS (correlation of -0.02); a schools' success in EP is very much predetermined by how academic their students are at intake.

Table 3: Pearson correlations between 5 A*-C, CVA, and EP in English and mathematics in 2010.

	5 A*-C	CVA	EP English	EP maths	KS2 APS
5 A*-C	1				
CVA	0.24	1			
EP English	0.85	0.36	1		
EP maths	0.89	0.29	0.77	1	
KS2 APS	0.87	-0.02	0.64	0.67	1

Notes: Number of schools = 3,056. 5 A*-C = Percentage of pupils with five or more GCSEs (or equivalent qualifications) at grade A* to C; CVA = Contextual value-added score; EP English = Percentage of pupils making expected progress in English; EP English = Percentage of pupils making expected progress in mathematics; KS2 APS = KS2 average point score.

Plotting schools' CVA ranks against their EP ranks (Figure 4) starkly reveals that many schools ranked high on EP are ranked low on CVA and vice versa. The two measure are clearly measuring very different things.

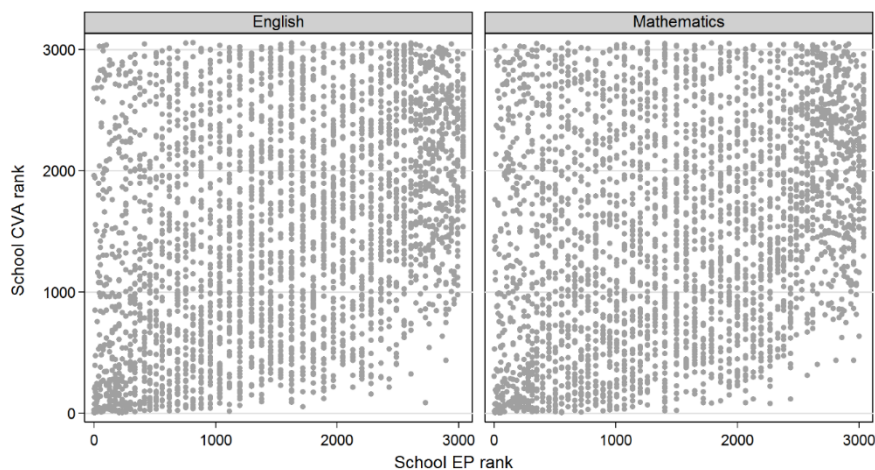


Figure 4: Scatterplot of school CVA ranks against EP ranks, based on 2010 school league table data, reported separately for English and mathematics. Higher ranks denote higher performances.

An interesting exercise is to consider how schools' ranks would likely change were the Government to revert back from EP to CVA and in particular, what types of schools would benefit or not by such a move. We plot the change in national rank against school mean KS2 APS (Figure 5). As expected, EP is strongly biased in favour of schools with high prior attaining intakes: schools with high prior attaining intakes would see very large drops in their national ranking were the Government to switch back from EP to CVA. The distinct cluster of schools which would particularly lose out with a return to CVA are 'grammar' schools, a small subset of around 160 schools which select pupils academically at intake and therefore have especially high school mean prior attainment. In grammar school areas, 'secondary modern' schools take the remaining pupils and so these schools therefore have especially low mean prior attainment. The point, however, is a more general one which is that CVA and EP are quite different school performance measures leading to substantially different

rankings which will be systematically biased in favour or against particular types of schools.

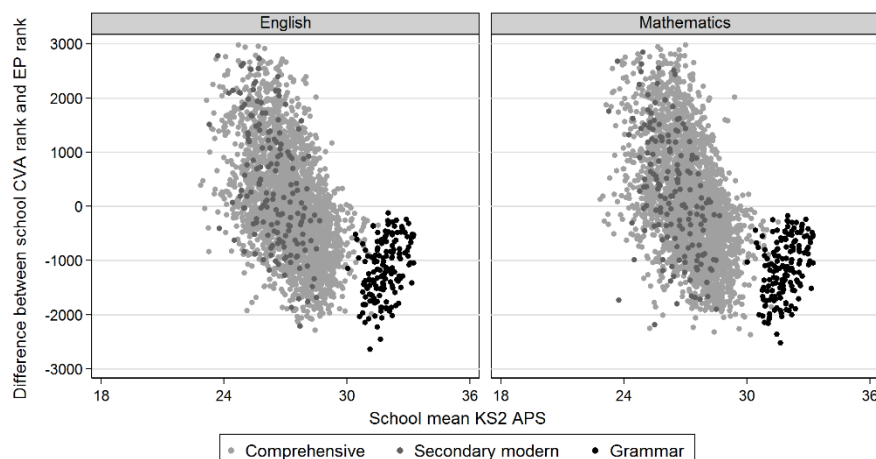


Figure 5: Difference between school CVA and EP ranks against school mean KS2 average point score, based on 2010 school league table data, reported separately for English and mathematics. KS2 levels map onto the KS2 point score scale as follows: [18,24) = KS2 level 3 (i.e., low prior attainers); [24,30) = KS2 level 4 (i.e., middle prior attainers); [30,36) = KS2 level 5 (i.e., high prior attainers).

Recall the Government's floor standards and that since 2011 a school is judged 'underperforming' if less than 40% of pupils achieve 5 A*-C, but is exempted if they are in the top half nationally in EP in both English and mathematics. In 2010, 464 schools (15%) were judged underperforming, a further 37 schools were excused due to outperforming the national median school in EP in English and mathematics. These figures drop to 303 schools (10%) with a further 198 schools now being excused when we use CVA in place of EP. What these simple statistics inform us is that the purported 'balancing role' played by EP is undermined by EP being much closer to a pure attainment measure rather than a pure progress measure. In contrast, CVA would have done a far better job in contextualizing schools' performances with respect to their different student compositions.

6 Progress 8

In 2016, the Government will withdraw EP replacing it with P8, a new value-added based measure derived from a multiple linear regression model, a simplified version of which can be written as

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + r_{ij}, \quad \hat{r}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} \hat{r}_{ij} \quad (3)$$

where

- y_{ij} denotes the GCSE score of pupil i ($i = 1, \dots, n_j$) in school j ($j = 1, \dots, J$)
- x_{ij} denotes their KS2 score
- r_{ij} denotes the pupil residual
- $\hat{\tau}_j$ denotes the predicted school value-added effect

P8 will adjust for a flexible function of pupil prior attainment (x_{ij}) and so should avoid the borderline effects and biases of EP. P8 scores ($\hat{\tau}_j$) will also once again be presented with 95% confidence intervals and therefore avoid that criticism of EP. However, P8 will continue to ignore school differences in the socioeconomic and demographic composition of their pupils.

P8 will also replace EP in the Government's floor standards. A school will now be judge underperforming if its pupils score on average half a grade lower than predicted and if this difference is statistically significant. Thus, the new floor standards will entirely be based on the new progress measure and statistical noise involved in these calculations will now be taken into account. These changes represent a substantial improvement on the Government's previous floor standards.

7 Conclusion

The UK Government's reasons for withdrawing CVA, a value-added based measure, are questionable. CVA's successor, EP, a value-table based measure, appears fundamentally flawed. In particular, EP perversely incentivises schools' efforts on borderline pupils, it is severely dependent on prior attainment, it ignores school differences in pupils' backgrounds, and it fails to communicate statistical uncertainty. CVA, while by no means perfect, largely avoided these pitfalls. P8 is conceptually a return to the value-added based approach of CVA and should therefore also avoid these pitfalls, however, it will continue to ignore pupils' socioeconomic and demographic backgrounds and we think this fundamentally problematic in terms of holding schools accountable.

References

1. Castellano, K. E. & Ho, A. D. (2013) A Practitioner's Guide to Growth Models (Council of Chief State School Officers).
2. Department for Education (2010) The importance of teaching: the schools white paper 2010 (Department for Education, London). URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/175429/CM-7980.pdf.
3. Goldstein, H. (2011). Multilevel Statistical Models, 4th edition. (Chichester, UK: Wiley).

4. Leckie, G., Charlton, C. & Goldstein, H. (2016) Communicating uncertainty in school value-added league tables. Centre for Multilevel Modelling, University of Bristol. URL: <http://www.cmm.bris.ac.uk/interactive/uncertainty/>.
5. Leckie, G. & Goldstein, H. (2009) The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172, 835-851.
6. Leckie, G. & Goldstein, H. (2011) Understanding uncertainty in school league tables. *Fiscal Studies*, 32, 207-224.